



ELSEVIER

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Growth and structure of Slovenia's scientific collaboration network

Matjaž Perc*

Department of Physics, Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, SI-2000 Maribor, Slovenia

ARTICLE INFO

Article history:

Received 17 March 2010

Received in revised form 18 April 2010

Accepted 22 April 2010

Keywords:

Scientific collaboration

Networks

Preferential attachment

Small worlds

Slovenia

ABSTRACT

We study the evolution of Slovenia's scientific collaboration network from 1960 till present with a yearly resolution. For each year the network was constructed from publication records of Slovene scientists, whereby two were connected if, up to the given year inclusive, they have coauthored at least one paper together. Starting with no more than 30 scientists with an average of 1.5 collaborators in the year 1960, the network to date consists of 7380 individuals that, on average, have 10.7 collaborators. We show that, in spite of the broad myriad of research fields covered, the networks form "small worlds" and that indeed the average path between any pair of scientists scales logarithmically with size after the largest component becomes large enough. Moreover, we show that the network growth is governed by near-linear preferential attachment, giving rise to a log-normal distribution of collaborators per author, and that the average starting year is roughly inversely proportional to the number of collaborators eventually acquired. Understandably, not all that became active early have till now gathered many collaborators. We also give results for the clustering coefficient and the diameter of the network over time, and compare our conclusions with those reported previously.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The structure of social networks is paramount for understanding the spread of knowledge, cultural traits, disease, as well as many other entities and attributes that can be associated with individuals living in groups or societies. As such it has been the subject of intense investigation, both theoretical as well as empirical, for at least half a century (Barabási, 2002; Christakis & Fowler, 2009; Wasserman & Faust, 1994; Watts, 1999). Primarily, and in many ways not really surprisingly, however, these investigations were in the domain of social rather than natural sciences. Probably best known in this context is the study by Milgram (1967), who studied how many steps it took, on average, to get a letter from a randomly selected person to a stockbroker in Boston, who was a friend of Milgram's. The result was six – a number that has since been reused outside of science for a number of purposes, one of the latest examples being the launch of SixDegrees.org seeking to exploit the "small-world phenomenon" for charitable purposes. A shortcoming of the study of Milgram, as well as that of many others conducted in a similar fashion, is that the size and structure of social networks mapped in such a direct and labor intensive way is rather small and receptive to bias. The advent of large-scale online portals made it possible to test the "six degrees of separation" hypothesis more thoroughly. Remarkably though, a study performed by Leskovec and Horvitz (2008), encompassing some 30 billion conversations from 240 million people, reported that the average path length among Microsoft Messenger users is 6.6. Although being closer to seven than six, the number is nevertheless in a strikingly good agreement with the result by Milgram obtained over 40 years ago.

* Tel.: +386 41 981 460.

E-mail address: matjaz.perc@uni-mb.si.URL: <http://www.matjazperc.com/>.

In natural sciences the attention to networks was sparked by works such as those of Watts and Strogatz (1998) and Barabási and Albert (1999), making ground-breaking advances with regard to our understanding of the “small-world phenomenon” and the emergence of scaling via growth and preferential attachment, respectively. The two works, along with subsequent refinements of the concepts they introduced (Barthélemy & Amaral, 1999; Newman, Moore, & Watts, 2000; Krapivsky, Redner, & Leyvraz, 2000; Dorogovtsev, Mendes, & Samukhin, 2000; Dorogovtsev & Mendes, 2000; Amaral, Scala, Barthélemy, & Stanley, 2000; Krapivsky & Redner, 2001; Krapivsky, Rodgers, & Redner, 2001), spawned an impressive number of studies on networks, as evidenced by the many reviews (Newman, 2000; Albert & Barabási, 2002; Dorogovtsev & Mendes, 2002; Newman, 2003; Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006; Dorogovtsev, Goltsev, & Mendes, 2008) and books (Dorogovtsev & Mendes, 2003; Pastor-Satorras & Vespignani, 2004; Newman, Watts, & Barabási, 2006; Barrat, Barthélemy, & Vespignani, 2008) dedicated either specifically to this field of research or its many interdisciplinary variations. This is all the more impressive since, at least within the hard sciences, prior to the late 1990s a paper on network theory is hard to come by (Newman, 2009). For a field this young the volume of insightful findings that have accumulated until now is something to be reckoned with. For example, the structure of networks has been found crucial for their resilience to error and attack (Albert, Jeong, & Barabási, 2000; Cohen, Erez, ben Avraham, & Havlin, 2000, 2001; Callaway, Newman, Strogatz, & Watts, 2000; Pietsch, 2006), for the fast availability of information within the world-wide-web (Albert, Jeong, & Barabási, 1999; Pastor-Satorras, Vázquez, & Vespignani, 2001), uninterrupted supply with electricity (Albert, Albert, & Nakarado, 2004), fast spread of epidemics and viral infections (Pastor-Satorras & Vespignani, 2002; Zanette & Kuperman, 2002; Barthélemy, Barrat, Pastor-Satorras, & Vespignani, 2004; Colizza, Barrat, Barthélemy, Valleron, & Vespignani, 2007), robust and near flawless reproduction of organisms (Hartwell, Hopfield, Leibler, & Murray, 1999), the evolution of cooperation (Santos & Pacheco, 2005; Szabó & Fátth, 2007; Perc, 2009) and coevolution (Gross & Blasius, 2008; Perc & Szolnoki, 2010), the dynamics of social systems (Castellano, Fortunato, & Loreto, 2009), and surely many other aspects of everyday life.

An interesting and potentially very revealing subset of complex networks are the social networks, of which scientific collaboration networks are a beautiful example (Newman, 2001d,b,c, 2004). Notably, for a social network to be representative for what it stands – an account of human interaction – a consistent definition of acquaintance is important. And while it may be challenging to *define* a friendship or an enemy in a consistent and precise manner (Moody, 2001; Moody & White, 2003), scientific collaboration is accurately documented in the final product and thus fairly straightforward to assess. Also amenable to a precise definition of connectedness are movie actors (Amaral et al., 2000), electric grids (Watts & Strogatz, 1998; Albert et al., 2004) and airports (Guimera, Mossa, Turttschi, & Amaral, 2005), for example, yet these are either approximations of social networks in that they don't really document real human contact or the level of acquaintance between people forming them is difficult to determine. As argued convincingly by Newman (2001d), considering scientific collaboration networks alleviates these problems to a large extent.

Here we study the evolution of a scientific collaboration network, namely that formed by Slovenia's scientists, from its very beginnings in the 1960s until the present time. Covering a time span of 50 years, the data are unique in that they provide an excellent testing ground for the “small-world” and preferential attachment hypotheses in growing social networks. We tackle these issues similarly as outlined in previous studies on growing scientific collaboration networks (Newman, 2001a; Jin, Girvan, & Newman, 2001; Barabási, Jeong, Néda, Ravasz, Schubert, & Vicsek, 2002; Jeong, Néda, & Barabási, 2003; Moody, 2004), where it has been reported, for example, that the growth is governed by linear or sublinear preferential attachment, and that as the networks grow their average degree increases while the average distance between individuals decreases. Evidences for strong clustering and models describing the growth of social networks have been presented in this context as well. Notably, for a set of different yet static scientific collaboration networks, Newman (2001d) has shown that the average distance between different authors scales logarithmically with size. We come to results that are in agreement with these earlier observations, but for a single growing scientific collaboration network. Moreover, we show that the observed near-linear preferential attachment rate translates into the expected log-normal degree distribution fairly accurately; a detail that was previously a source of some discrepancy not just in the context of scientific collaboration networks (Jeong et al., 2003; Redner, 2005). In the continuation we first give information on the raw data and network construction, while the results are presented and summarized in Sections 3 and 4, respectively.

2. Preliminaries

Slovenia is a small country located at the heart of Europe with a population of two million.¹ It has a well-documented research history, which is made possible by SICRIS – Slovenia's Current Research Information System² – hosting up-to-date publication records of all Slovene scientists. At present, there are 30,630 registered, including young and non-active scientists as well as laboratory personnel, which boils down the initial number to 8402 of those that are truly active research-wise or have been so in the past. By this we mean those that, to date, have at least one bibliographic unit indexed by the Web of Science. This criterium may be somewhat stringent, but it is the only one we could apply consistently. Moreover, since the publication data contain records not just of research but also of professional work and many other activities not necessarily concerning research, it is important to define a threshold for when two scientists are considered connected. Having given

¹ The official Web page of Slovenia is accessible via: <http://www.slovenia.si/>.

² The SICRIS Web page is accessible via: <http://sicris.izum.si/>.

an interview together can hardly be compared to writing a joint research paper. In accordance with previous studies, which by default considered databases containing almost exclusively research papers and where therefore the setting of such a threshold was not necessary, we here consider two scientists as being connected if, up to the given year inclusive, they have coauthored at least one research or review paper together. A final factor slightly affecting the network size is the necessity of identifying authors on a full first and last name basis. Identification based on the last name and first initial only ($< 1\%$ of Slovene population has more than one initial) has already been shown to downsize scientific collaboration networks considerably (Newman, 2001d). Here this problem is additionally amplified by the fact that certain last names are very frequent, for example because a family is traditionally involved in research over many generations, but also because in any given country certain last names are far more common than others. In our case this introduces an unacceptably large bias, grouping too many different individuals together and making them appear as a single author. On the other hand, requiring full first and last name agreement for identification practically never confuses two different authors for one, but is likely to miss out a few links if authors do not use their first names consistently. Although all publication records in SICRIS are always given with full first and last names (irrespective of whether the actual paper features just the initials of the first name), sometimes authors are evidenced in SICRIS under a given first name but then use an abbreviated (e.g. "Alojzij" may become "Alojz") or internationalized (e.g. "Aleksander" may become "Alex") version thereof, and SICRIS is not always consistent in eliminating such discrepancies. Challenging for parsing are also the non-ASCII characters that are quite common in Slovene names, such as č, ć, š and ž, but this can be solved with a high success rate by converting them all to standard ASCII counterparts via a unique-enough rule. Taking these considerations into account, we start with 30 scientists with an average of 1.5 collaborators in the year 1960 and end with 7380 scientists with an average of 10.7 collaborators in 2010 with a yearly time resolution in-between, assured that the error margin is highly unlikely to affect the presented results in any perceivable way. To date the 7380 scientists forming the network in 2010 have written 76,194 papers, altogether having 735,619 citations assigned to them. A detailed statistical analysis of individual scientific indicators, which is an interesting and vibrant subject on its own (Egghe & Rousseau, 1990; Hirsch, 2005; Egghe, 2006, 2008; Schreiber, 2008a,b; Rousseau & Ye, 2008; Schreiber, 2010) is, however, presented in a separate study (Perc, 2010).³

3. Results

We start with presenting the evolution of fundamental statistical properties of the scientific collaboration network in Fig. 1. The inset of panel A shows how the network size increases with time. On a semi-log scale we could attempt making a linear fit for at least some time windows, yet the obtained exponents would not be very meaningful given that a slightly negative curvature is obviously present across the majority of the plot. We therefore satisfy ourselves with the conclusion that the network growth is almost exponential in time, which is certainly a result worth some admiration. Note that exponential growth is usually associated with online networks such as Delicious or Yahoo! Answers (Leskovec, Backstrom, Kumar, & Tomkins, 2008), but could hardly be expected for a scientific collaboration network. This result can be taken as evidence that science is well taken care of in Slovenia, but also that it is not yet at capacity in terms of the number of scientists that can be sustained. The main graph in panel A, on the other hand, shows the fraction of the largest component with respect to the total network size in the pertaining year. Initially the network was quite fragmented with the largest component occupying less than 40% of its overall size. In good 20 years, however, the largest component started exceeding 2/3 of the network. Currently it occupies 98.7%, meaning that the vast majority of Slovenia's scientists is connected with one another via some number of intermediate acquaintances. Those not belonging to this giant component typically form small isolated groups that don't crucially affect the structure of the network. As discussed already by Newman (2001d), this is a desirable feature since it signalizes that science as a whole is a product of joint and strongly interdisciplinary rather than isolated efforts. This, in turn, can be seen as an important driving force behind the ever faster scientific progress and innovation.

The emergence of the giant component brings with it some noticeable changes in the way certain quantifiers of the network structure evolve over time. In order to highlight this transition, each panel of Fig. 1 has a horizontal axis break giving visual priority to the 1985–2010 period, during which the giant component occupies more than 70% of the network. The main graph in panel B shows that the clustering coefficient, calculated as three times the number of triangles divided by the number of connected triples of vertices [see e.g. Newman (2003)], starts decreasing monotonously, while simultaneously the average number of collaborators per author (average degree of the network) assumes a steady exponential increase (see inset). Prior to 1985 the trends are less clear and fluctuating, but qualitatively similar. Although the clustering coefficient is decreasing over the years, with around 20% chance of two scientist collaborating if both have done so with a third scientist in 2010, it can still be concluded that clustering is an inherent property of the studied scientific collaboration network. Altogether these results are in good agreement with earlier findings obtained for larger scientific collaboration networks but over a shorter time span (Barabási et al., 2002). Notably, large increments in the level of collaboration over an even longer time span then considered here have been reported also by Grossman (2002) for a scientific collaboration network of mathematicians (Grossman & Ion, 1995).

Most impressive, however, is the shift in the dependence of the mean distance between authors and the diameter of the network, as depicted in panel C of Fig. 1 (main graph and inset). The year 1985 constitutes a tipping point, following

³ Tables of scientific indicators for Slovene researchers are accessible via: <http://www.matjazperc.com/sicris/stats.html>.

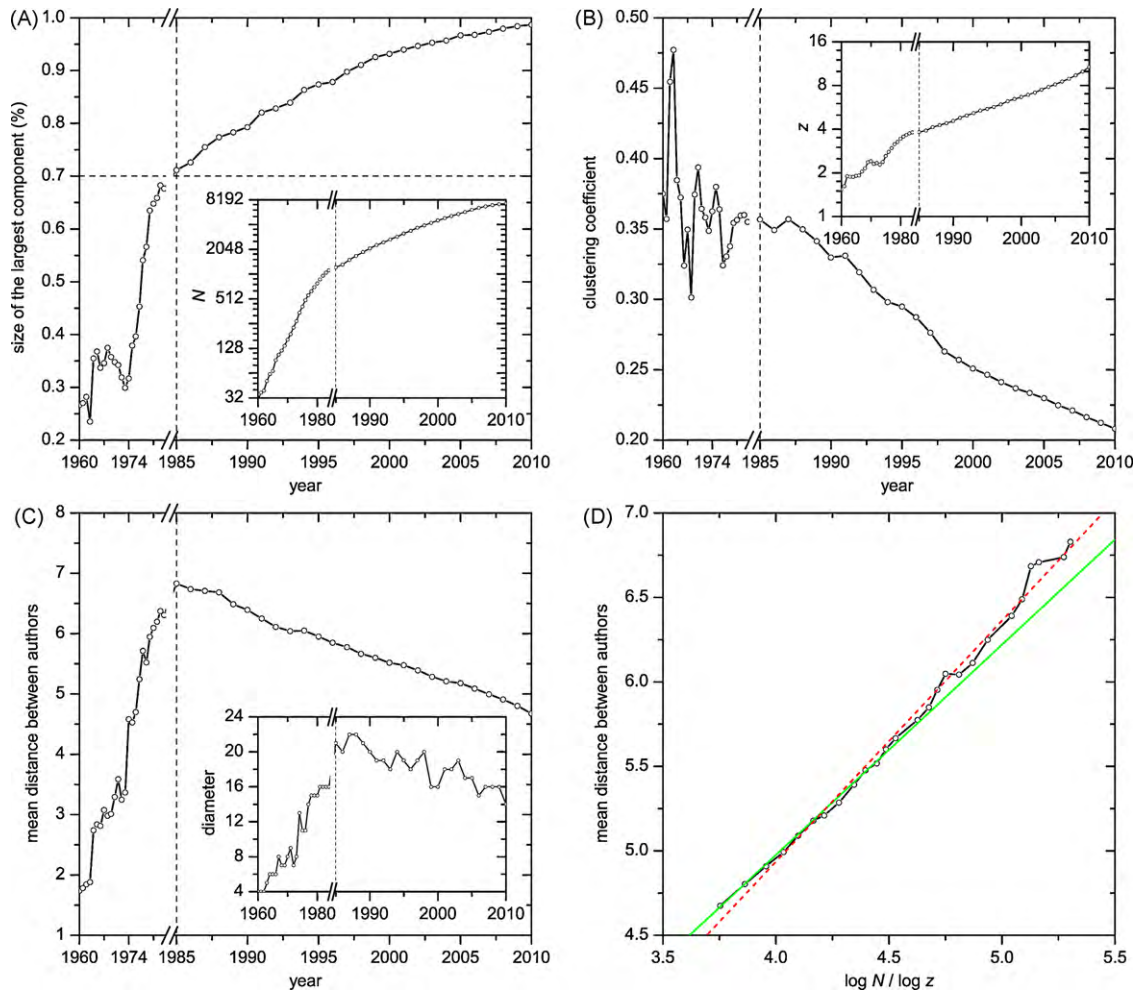


Fig. 1. Fundamental statistical properties of the scientific collaboration network over time. All panels have a horizontal axis break, marked by the vertical dashed line, such that the period from 1960 to 1984 occupies 20% (30% in insets) of the total axis span, while the period from 1985 to 2010 occupies the remainder. The axis break is motivated by the emergence of the logarithmic variation of the mean distance with size (panel D) and the largest component exceeding 70% of the network (panel A). This also coincides with the settling of other network properties into predictable paths past the year 1985. (A) Size of the largest component in percentage of the total network size N . Dashed horizontal line marks the 70% borderline, which is exceeded in 1985. The inset shows network growth, starting with $N = 30$ scientists in 1960 and ending with $N = 7380$ in 2010. Network growth is almost exponential in time, with the curve having a slightly negative curvature on a semi-log scale. (B) Clustering coefficient and the average number of collaborators z (inset) over time. Past the year 1985 z increases exponentially. (C) Mean distance between authors and the diameter (inset) over time. The tipping point is the year 1985, following which both quantities gain a downward momentum. (D) Mean distance between authors past 1985 in dependence on $\log N / \log z$. Dashed red line is the best linear fit to the data while the solid green line is the best linear fit going also through the origin. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

which the upward trend prior is replaced by a steady downward trend. However, since the diameter is obviously more prone to fluctuations than the mean distance, the tipping point in the inset is expressed less accurately, but still clearly. It is remarkable that in spite of the broad myriad of research fields covered by Slovenia's scientists, the networks still form "small worlds" in that the path between any given pair (provided it exists) leads across only a few – on average less than five but definitely not more than 14 in 2010 – intermediate acquaintances. It is common to compare these values with what would be obtained on a random graph having the same size and average degree (Bollobás, 1985). Indeed, random graphs can be considered as the simplest of models of social networks. Panel D of Fig. 1 shows how the average distance varies in dependence on $\log N / \log z$ (exactly the variation of the mean distance observed on random graphs) within the 1985–2010 period, along with the best linear fit (dashed red line) and the best straight line fit going also through the origin (solid green line). Evidently, there is a considerable overlap with the logarithmic variation that can be observed for random graphs, especially past the year 1995, where the points fall almost perfectly onto the solid green line, thus confirming the existence of the "small-world phenomenon" in the studied scientific collaboration network. Notably, this result is interesting on its own as an empirical demonstration of logarithmic variation with size in a real growing social network.

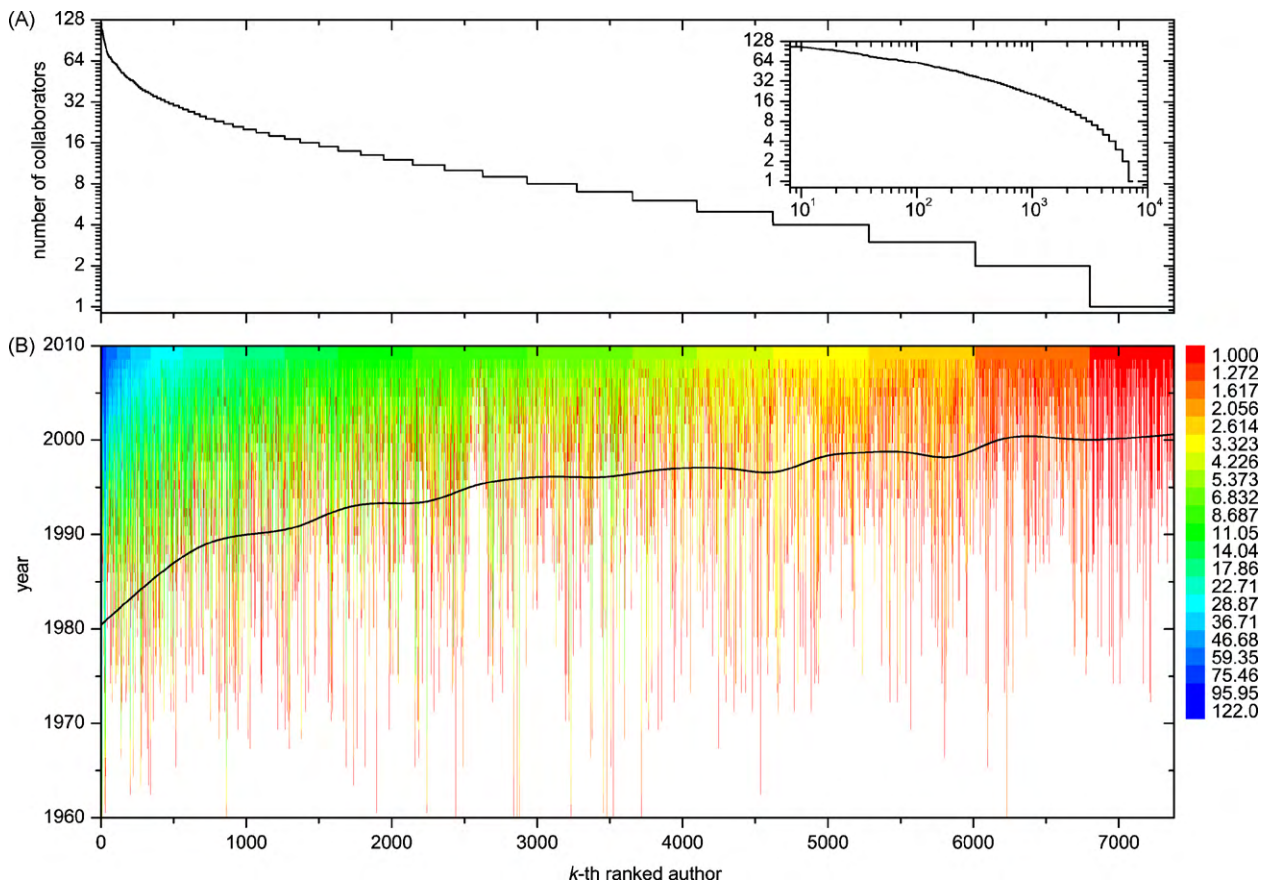


Fig. 2. (A) Zipf plot of the number of collaborators versus the k -th ranked author on a semi-log (inset shows log–log) scale, as obtained for the year 2010. (B) Color-coded evolution of individual collaborator numbers leading to the Zipf plot in panel A. Each vertical line refers to a single author, whereby authors are ranked according to the total number of collaborators they have in the year 2010 (see panel A). The start of each vertical line corresponds to the year the pertaining author received his/her first collaborator, i.e. when s/he became active and thus a part of the collaboration network. The coloring denotes how the number of collaborators of each author increased over the years (from the time of becoming active till 2010), according to the color bar on the right. Not all authors that became active early eventually gathered many collaborators, although there is a fairly clear trend towards an earlier start when approaching the 1st ranked author. This is evidenced by the solid black line, depicting the average starting years that were determined within non-overlapping windows each containing 300 consecutively ranked authors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

In addition to the above presented results, the data enable testing the preferential attachment hypothesis in growing social networks. Growth and preferential attachment have been proposed by [Barabási and Albert \(1999\)](#) as the two crucial factors for the emergence of scaling in random networks. Notably, synonymous to preferential attachment are the terms cumulative advantage ([de Solla Price, 1965, 1976](#)) and the Matthew effect ([Merton, 1968](#)), which were proposed in the 60s but attracted considerably less attention. It is also interesting to note that the general evolutionary theory of information production processes can be applied successfully on evolving complex networks ([Egghe, 2007](#)), and that in this context the so-called “success-breeds-success” principle ([Egghe & Rousseau, 1995, 2006](#)) was used synonymously. In essence, however, these terms all describe that the more one has (of something), the more likely it is one will eventually gain even more; and conversely, that the ones that have very little are unlikely to recover. Translating this to growing scientific collaboration networks, it means that the more collaborators one has at a given time, the more likely it is one will attract new scholars in the future. On the other hand, if one has only a few collaborators, it is less likely s/he will establish new acquaintances. Indeed, this seems like a very reasonable proposition.

Panel B of [Fig. 2](#) shows how the number of collaborators increases over the years for every scientist considered in this study. Individuals are thereby enumerated according to their rank, as determined by the total number of collaborators they have in the year 2010. Number one is the scientist that in 2010 has the highest degree (equalling 122), author number two has the second-largest degree (equalling 119), and so on. The upper-most color stripe in [Fig. 2B](#) (for the year 2010) thus gives the Zipf plot ([Zipf, 1949](#)) of the number of collaborators in the year 2010. Since the Zipf plot can give vital clues on the expected distribution of the examined quantity, we show it in panel A of [Fig. 2](#) separately on a semi-log and log–log (inset) scale. From the informetrics perspective it is noteworthy that the Lotka law, describing for example the frequency of publication by authors, is often equivalent to the Zipf law in that it also assumes an inverse square form [see e.g. [Egghe](#)

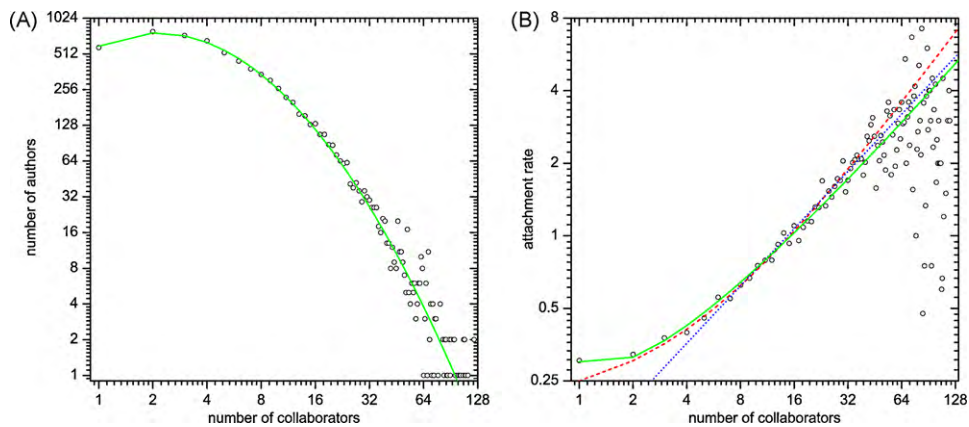


Fig. 3. (A) Distribution of the number of collaborators y on a log–log scale. Solid green line is the log-normal fit $P(y) \sim \exp[a \ln y - b(\ln y)^2]$ of the data, with $a = 0.69$ and $b = 0.46$. (B) Attachment rate $A(y) = \Delta y / y$ versus the number of collaborators on a log–log scale. Depicted is the average over 50 years, where y was determined in the year x and Δy in the year $x + 1$ for $x = 1960, 1961, \dots, 2009$. Solid green line is the nearly linear fit $A(y) \sim y / (1 + c \ln y)$, with $c = 1.32$. For comparison, dotted blue line is the sublinear fit $A(y) \sim y^g$, with $g = 0.79$, and the dashed red line is the linear fit $A(y) \sim hy$, with $h = 0.053$. While sublinear attachment rates give rise to stretched exponential distributions (Krapivsky et al., 2000), log-normal distributions arise from nearly linear preferential attachment (Redner, 2005), thus making results presented in panels A and B in good agreement with one another. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

(2005) for an interesting treatment]. Had the Zipf plot on the log–log scale a linear outlay with the slope β , this would be equivalent to a power-law distribution of the number of collaborators y of the form $P(y) \sim y^{-\gamma}$, with $\gamma = 1 + 1/\beta$. Furthermore, this would then imply linear preferential attachment governing the growth of the studied network. Yet this is not the case since the Zipf plot on the log–log scale has an obvious negative curvature. The semi-log scale is similarly inconclusive, thus hinting towards an interesting degree distribution and the underlying attachment rate.

Before determining this, however, it is worth examining how the year of becoming active translates to the number of collaborators one eventually acquires over the years. Although becoming active in the sense of writing the first research paper is not necessarily identical to becoming active in the sense of acquiring at least one collaborator (for the obvious reason that one can write a research paper as a sole author), this discrepancy is small enough to be negligible. The color map presented in panel B of Fig. 2 shows that the sooner one started acquiring collaborators, the more likely it is one has a lot of them in the year 2010. In other words, the age of a node is roughly inversely proportional to its degree, which is characteristic for networks whose growth is governed by preferential attachment [see e.g. Ben-Naim and Krapivsky (2009) for a recent treatment]. However, for growing scientific collaboration networks this conclusion is obviously challenged for several reasons. Note that a sliding non-overlapping window of 300 consecutively ranked scientists yields the average starting year as depicted by the black solid line in panel B, but individual discrepancies to this are far from uncommon. An obvious reason contributing to the deviation is that not all people are equally fond of making new acquaintances. This is true for life in general, and probably even more so for scientific collaboration. Moreover, a scientist may have gone dormant in the course of time, thus stopping the acquirement of new collaborators. Nevertheless, the statement that older scientists are more likely to have more collaborator than younger scientists is certainly reasonable and valid, and results presented in panel B of Fig. 2 support such a conclusion to a large extent.

Turning now to testing the preferential attachment hypothesis, we first show in panel A of Fig. 3 the degree distribution of the network in the year 2010, i.e. the number of authors with a given number of collaborators. On a log–log scale the points exhibit, similarly as the Zipf plot in Fig. 2, a negative curvature across the whole span of the number of collaborators. Accordingly, a power-law fit, even with an exponential cutoff, cannot describe these data accurately [see e.g. Clauset, Shalizi, and Newman (2009)]. Indeed, it turns out that a log-normal form, as given in the caption of Fig. 3 is most fitting, as evidenced by the solid green line. While log-normal forms are typically associated with random multiplicative processes [see e.g. Krapivsky and Redner (2007)], recently Redner used it for fitting the cumulative citation distribution of the Physical Review over the past 110 years. There it was also pointed out that such distributions may arise from near-linear preferential attachment. To test this for our social network we have determined the average attachment rate $A(y)$, giving the likelihood that an author with y collaborators in a given year will have Δy more collaborators in the next year, as specified in the caption of Fig. 3. This way of measuring the preferential attachment in evolving networks was proposed by Jeong et al. (2003), and used also for the assessment of preferential attachment in a growing scientific collaboration network, albeit over a shorter time span than considered here, by Barabási et al. (2002). Results presented in the panel B of Fig. 3 show that the data can be fitted best by the near-linear form $A(y) \sim y / (1 + c \ln y)$ (solid green line). For comparison, a linear fit (dashed red line) and a “traditional” sublinear fit (dotted blue line) are shown as well. From this we can conclude that the growth of Slovenia’s scientific collaboration network is governed by near-linear preferential attachment, which translates fairly accurately into the expected log-normal distribution of collaborators per author. Indeed, near-linear attachment rates

of the form $A(y) \sim y/(1 + c \ln y)$ yield log-normal distributions as depicted in panel A of Fig. 3 (Redner, 2005), thus giving a coherent ending to our study.

4. Summary

In sum, we have studied the growth and structure of Slovenia's scientific collaboration network over the past 50 years, focusing specifically on testing the “small-world” and preferential attachment hypotheses, but analyzing also other aspects of the network structure in detail. We have shown that there exists a tipping point in time after which the mean distance between authors and the diameter start decreasing. Accompanied by an exponential increase of the average number of collaborators per author and a near-exponential growth of the network over time, this gives rise to a logarithmic variation of the mean distance with size, much in agreement with what would be observed for growing random networks. Notably, the emergence of the tipping point coincides with the largest component exceeding 70% of the network size, but roughly also with the breakup of Yugoslavia and the subsequent downfall of the socialist regime. It is unclear whether the latter fact contributed to the emergence of the transition, but the time-wise correlation gives some room for speculations. Turning back to the network growth and structure, we have shown that the clustering coefficient decreases in time, but by hovering quite comfortably over 0.2 in the year 2010, it can be concluded that the clustering is still fairly strong. Finally, we have shown that the growth of the network is governed by near-linear preferential attachment that translates fairly accurately into the expected degree distribution, and that the year of becoming active scales inversely with the number of collaborators thus far acquired. Although deviations from linear preferential attachment have been reported before (Newman, 2001a; Barabási et al., 2002; Jeong et al., 2003), also outside the realm of scientific collaboration networks (Redner, 2005; Capocci et al., 2006), an agreement between the attachment rate and the resulting distribution derived from empirical data seems to be quite rare. Altogether, however, the presented results are in agreement with previous findings, supporting the conclusion that the growth of social networks is governed by preferential attachment, and that the resulting network structure has properties characteristic for “small worlds”.

Acknowledgments

Matjaž Perc thanks Matej Horvat from Hermes SofLab for illuminating lessons on socket programming and automated information retrieval from the Internet. Financial support from the Slovenian Research Agency (grant Z1-2032) is gratefully acknowledged as well.

References

- Albert, R., Albert, I., & Nakarado, G. L. (2004). Structural vulnerability of the north american power grid. *Physical Review E*, 69, 025103(R).
- Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47.
- Albert, R., Jeong, H., & Barabási, A.-L. (1999). The diameter of the world-wide web. *Nature*, 401, 130.
- Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406, 378.
- Amaral, L. A. N., Scala, A., Barthélemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11149.
- Barabási, A.-L. (2002). *Linked: The New Science of Networks*. Cambridge: Perseus.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509.
- Barabási, A.-L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A*, 311, 590.
- Barrat, A., Barthélemy, M., & Vespignani, A. (2008). *Dynamical Processes in complex networks*. Cambridge: Cambridge University Press.
- Barthélemy, M., & Amaral, L. A. N. (1999). Small-world networks: Evidence for a crossover picture. *Physical Review Letters*, 82, 3180.
- Barthélemy, M., Barrat, A., Pastor-Satorras, R., & Vespignani, A. (2004). Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters*, 92, 178701.
- Ben-Naim, E., & Krapivsky, P. L. (2009). Stratification in the preferential attachment network. *Journal of Physics A: Mathematical Theory*, 42, 475001.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424, 175.
- Bollobás, B. (1985). *Random graphs*. New York: Academic Press.
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H., & Watts, D. J. (2000). Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85, 5468.
- Capocci, A., Servedio, V. D. P., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., & Caldarelli, G. (2006). Preferential attachment in the growth of social networks: The internet encyclopedia wikipedia. *Physical Review E*, 74, 036116.
- Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81, 591.
- Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. New York: Little, Brown and Co.
- Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51, 661.
- Cohen, R., Erez, K., ben Avraham, D., & Havlin, S. (2000). Resilience of the internet to random breakdowns. *Physical Review Letters*, 85, 4626.
- Cohen, R., Erez, K., ben Avraham, D., & Havlin, S. (2001). Breakdown of the internet under intentional attack. *Physical Review Letters*, 86, 3682.
- Colizza, V., Barrat, A., Barthélemy, M., Valleron, A.-J., & Vespignani, A. (2007). Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Medicine*, 4, e13.
- de Solla Price, D. J. (1965). Networks of scientific papers. *Science*, 149, 510.
- de Solla Price, D. J. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27, 292.
- Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2008). Critical phenomena in complex networks. *Reviews of Modern Physics*, 80, 1275.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2000). Scaling behaviour of developing and decaying networks. *Europhysics Letters*, 52, 33.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, 51, 1079.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2003). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford: Oxford University Press.
- Dorogovtsev, S. N., Mendes, J. F. F., & Samukhin, A. N. (2000). Structure of growing networks with preferential linking. *Physical Review Letters*, 85, 4633.
- Egghe, L. (2005). Relations between the continuous and the discrete Lotka power function. *Journal of the American Society for Information Science*, 56, 664.

- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69, 131.
- Egghe, L. (2007). General evolutionary theory of information production processes and applications to the evolution of networks. *Journal of Informetrics*, 1, 115.
- Egghe, L. (2008). The influence of transformations on the h-index and the g-index. *Journal of the American Society for Information Science*, 59, 1304.
- Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Amsterdam: Elsevier.
- Egghe, L., & Rousseau, R. (1995). Generalized success-breeds-success principle leading to time-dependent informetric distributions. *Journal of the American Society for Information Science*, 46, 426.
- Egghe, L., & Rousseau, R. (2006). Systems without low-productive sources. *Information Processing & Management*, 42, 1428.
- Gross, T., & Blasius, B. (2008). Adaptive coevolutionary networks: A review. *Journal of the Royal Society, Interface*, 5, 259.
- Grossman, J. W. (2002). The evolution of the mathematical research collaboration graph. *Congressus Numerantium*, 158, 202.
- Grossman, J. W., & Ion, P. D. F. (1995). On a portion of the well-known collaboration graph. *Congressus Numerantium*, 108, 129.
- Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 7794.
- Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, 402, 47.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 16569.
- Jeong, H., Néda, Z., & Barabási, A.-L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61, 567.
- Jin, E. M., Girvan, M., & Newman, M. E. J. (2001a). Structure of growing social networks. *Physical Review E*, 64, 046132.
- Krapivsky, P. L., & Redner, S. (2001). Organization of growing random networks. *Physical Review E*, 63, 066123.
- Krapivsky, P. L., & Redner, S. (2007). Smoothing a rock by chipping. *Physical Review E*, 75, 031119.
- Krapivsky, P. L., Redner, S., & Leyvraz, F. (2000). Connectivity of growing random networks. *Physical Review Letters*, 85, 4629.
- Krapivsky, P. L., Rodgers, G. J., & Redner, S. (2001). Degree distributions of growing random networks. *Physical Review Letters*, 86, 5401.
- Leskovec, J., Backstrom, L., Kumar, R., & Tomkins, A. (2008). *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* ACM, New York, (p. 462).
- Leskovec, J., & Horvitz, E. (2008). *Proceeding of the 17th international conference on World Wide Web* ACM, New York, (p. 915).
- Merton, R. K. (1968). The matthew effect in science. *Science*, 159, 56.
- Milgram, S. (1967). The small world problem. *Psychology Today*, 2, 60.
- Moody, J. (2001). Race, school integration, and friendship segregation in America. *American Journal of Sociology*, 107, 679.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69, 213.
- Moody, J., & White, D. R. (2003). Social cohesion and embeddedness. *American Sociological Review*, 68, 103.
- Newman, M. E. J. (2000). Models of the small world. *Journal of Statistical Physics*, 101, 819.
- Newman, M. E. J. (2001a). Clustering and preferential attachment in growing networks. *Physical Review E*, 64, 025102(R).
- Newman, M. E. J. (2001b). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64, 016131.
- Newman, M. E. J. (2001c). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64, 016132.
- Newman, M. E. J. (2001d). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 404.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45, 167.
- Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5200.
- Newman, M. E. J. (2009). The first-mover advantage in scientific publication. *EPL*, 86, 68001.
- Newman, M. E. J., Moore, C., & Watts, D. J. (2000). Mean-field solution of the small-world network model. *Physical Review Letters*, 84, 3201.
- Newman, M. E. J., Watts, D. J., & Barabási, A.-L. (2006). *The structure and dynamics of networks*. Princeton: Princeton University Press.
- Pastor-Satorras, R., Vázquez, A., & Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical Review Letters*, 87, 258701.
- Pastor-Satorras, R., & Vespignani, A. (2002). Immunization of complex networks. *Physical Review E*, 65, 036104.
- Pastor-Satorras, R., & Vespignani, A. (2004). *Evolution and structure of the Internet: A statistical physics approach*. Cambridge: Cambridge University Press.
- Perc, M. (2009). Evolution of cooperation on scale-free networks subject to error and attack. *New Journal of Physics*, 11, 033027.
- Perc, M. (2010). Zipf's law and log-normal distributions in measures of scientific output across fields and institutions: 40 years of Slovenia's research as an example. *Journal of Informetrics*. doi:10.1016/j.joi.2010.03.001.
- Perc, M., & Szolnoki, A. (2010). Coevolutionary games—A mini review. *BioSystems*, 99, 109.
- Pietsch, W. (2006). Derivation of the percolation threshold for the network model of Barabási and Albert. *Physical Review E*, 73, 066112.
- Redner, S. (2005). Citation statistics from 110 years of physical review. *Physics Today*, 58, 49.
- Rousseau, R., & Ye, F. Y. (2008). A proposal for a dynamic h-type index. *Journal of the American Society for Information Science*, 59, 1853.
- Santos, F. C., & Pacheco, J. M. (2005). Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters*, 95, 098104.
- Schreiber, M. (2008a). A modification of the h-index: The h_m -index accounts for multi-authored manuscripts. *Journal of Informetrics*, 2, 211.
- Schreiber, M. (2008b). To share the fame in a fair way, h_m modifies h for multi-authored manuscripts. *New Journal of Physics*, 10, 040201.
- Schreiber, M. (2010). Revisiting the g-index: The average number of citations in the g-core. *Journal of the American Society for Information Science*, 61, 169.
- Szabó, G., & Fáth, G. (2007). Evolutionary games on graphs. *Physics Reports*, 446, 97.
- Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge: Cambridge University Press.
- Watts, D. J. (1999). *Small worlds*. Princeton: Princeton University Press.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393, 440.
- Zanette, D. H., & Kuperman, M. (2002). Effects of immunization on small-world epidemics. *Physica A*, 309, 445.
- Zipf, G. K. (1949). *Human behavior and the principle of least-effort*. Reading: Addison-Wesley.